

# Grid Computing: Technology for Scientific Collaboration



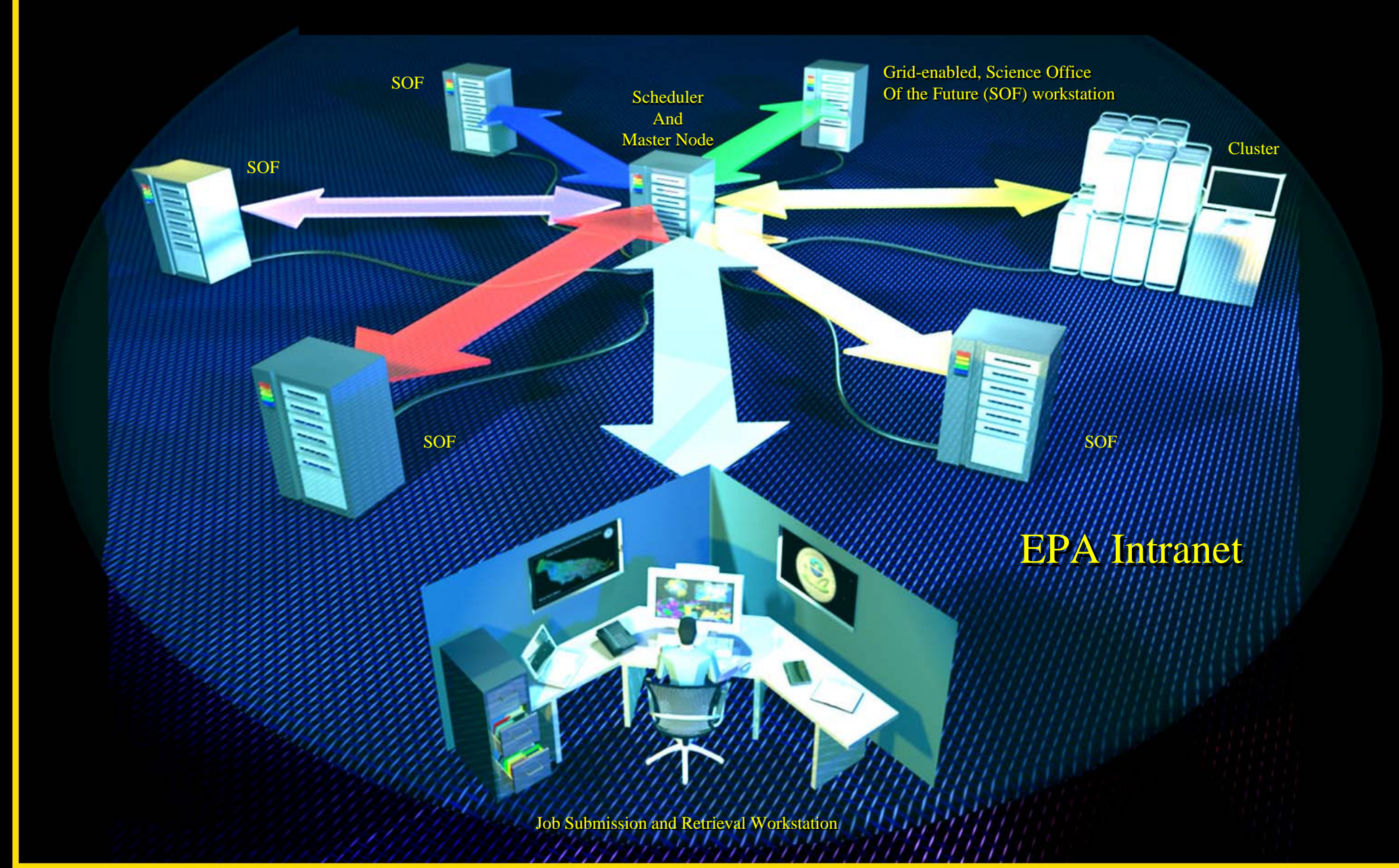
Author(s): Ravi Nair†, Gary Walter‡ (Science Coordinator)

Affiliation(s): †U.S. EPA/Office of Environmental Information (OEI)/Office of Technology Operations and Planning (OTOP)/National Computing Center (NCC); ‡ U.S. EPA/Office of Research and Development/National Exposure Research Laboratory/Atmospheric Sciences Modeling Division

EPA's Office of Environmental Information/Office of Technology Operations and Planning/NCC has a series of pilot projects under way to explore the feasibility of using emerging grid computing technology to enable EPA scientists and their collaborators to cost-effectively share their scientific computer and data resources. There is widespread interest in grid computing to foster such collaboration, both within the US and internationally. Many of the EPA's traditional science collaborators, such as the National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), and National Center for Health Statistics, have their own pilot projects to tap into the potential that this technology has to offer.

Grid computing offers a model for solving massive computational problems by allowing a user to tap into the unused resources (CPU cycles and/or disk storage) of disparate computers distributed across the network. The collection of unused resources that are available to the user is presented as a single system image. Grid computing also enables transparent access to data on storage resources that are part of the grid. Grid computing involves a set of rules for sharing heterogeneous, networked resources (different computing platforms, hardware/software architectures, and data) in geographically dispersed locations and belonging to different administrative domains using open standards. Grid technologies advance environmental research and development by allowing organizations to manage data locally and provide data globally.

## Compute Grid Pilot



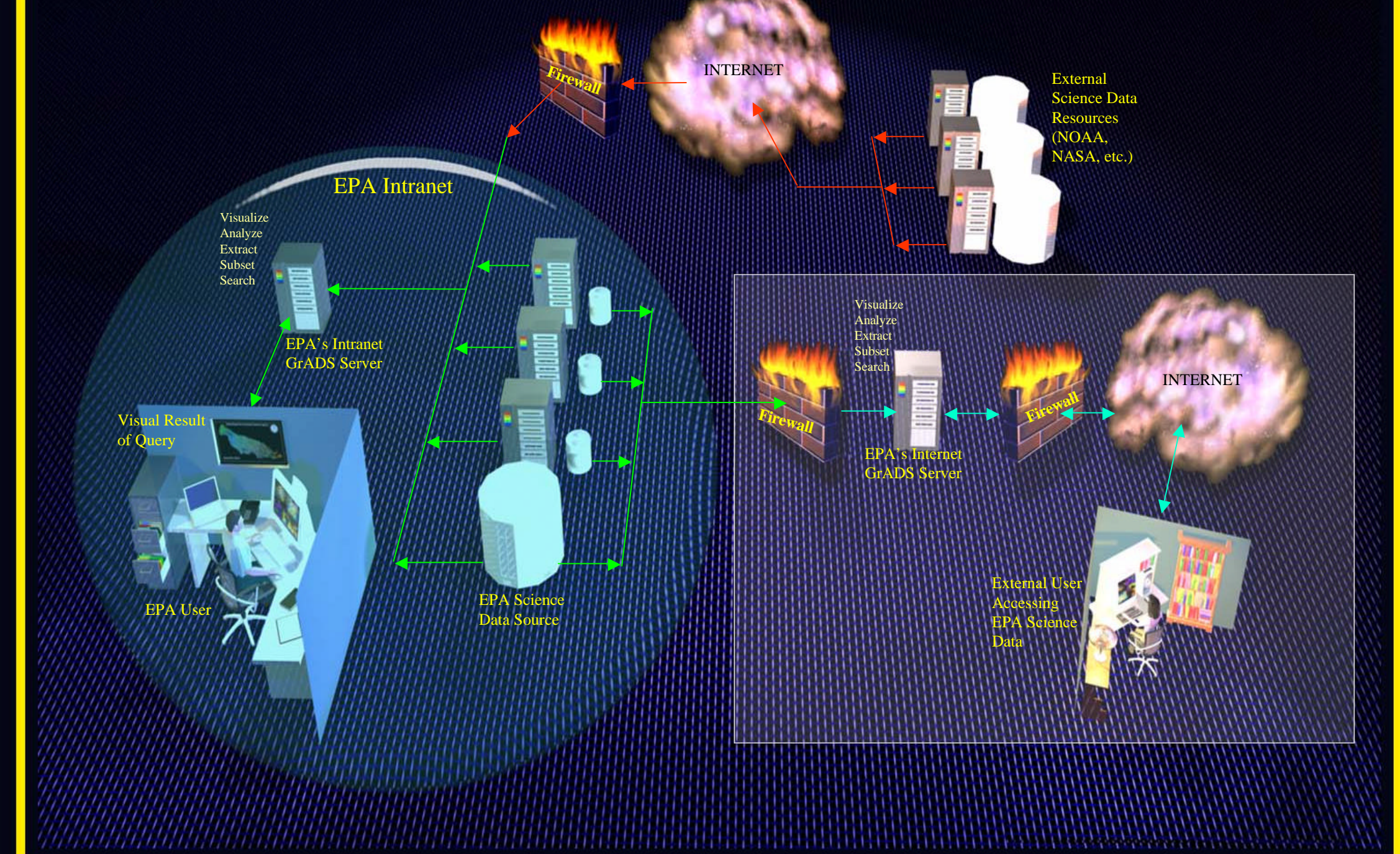
## Compute Grid Pilot

- Goal: Access unused cycles available within the organization to run computationally-intensive jobs, including parallel applications.
- Status:
  - Grid-enabled eight Scientific Office of the Future (SOF) workstations and an Aspen Systems 32-batch processor cluster in EPA's RTP Campus.
  - Successfully ran ORD/NERL's Environmental Fluid Dynamics Code (EFDC) and over fifteen hundred jobs of NERL's Air Pollution Exposure (APEX) Model.
  - Will begin runs of 8-processor-scaled versions of the Community Multiscale Air Quality (CMAQ) Model by end of May.

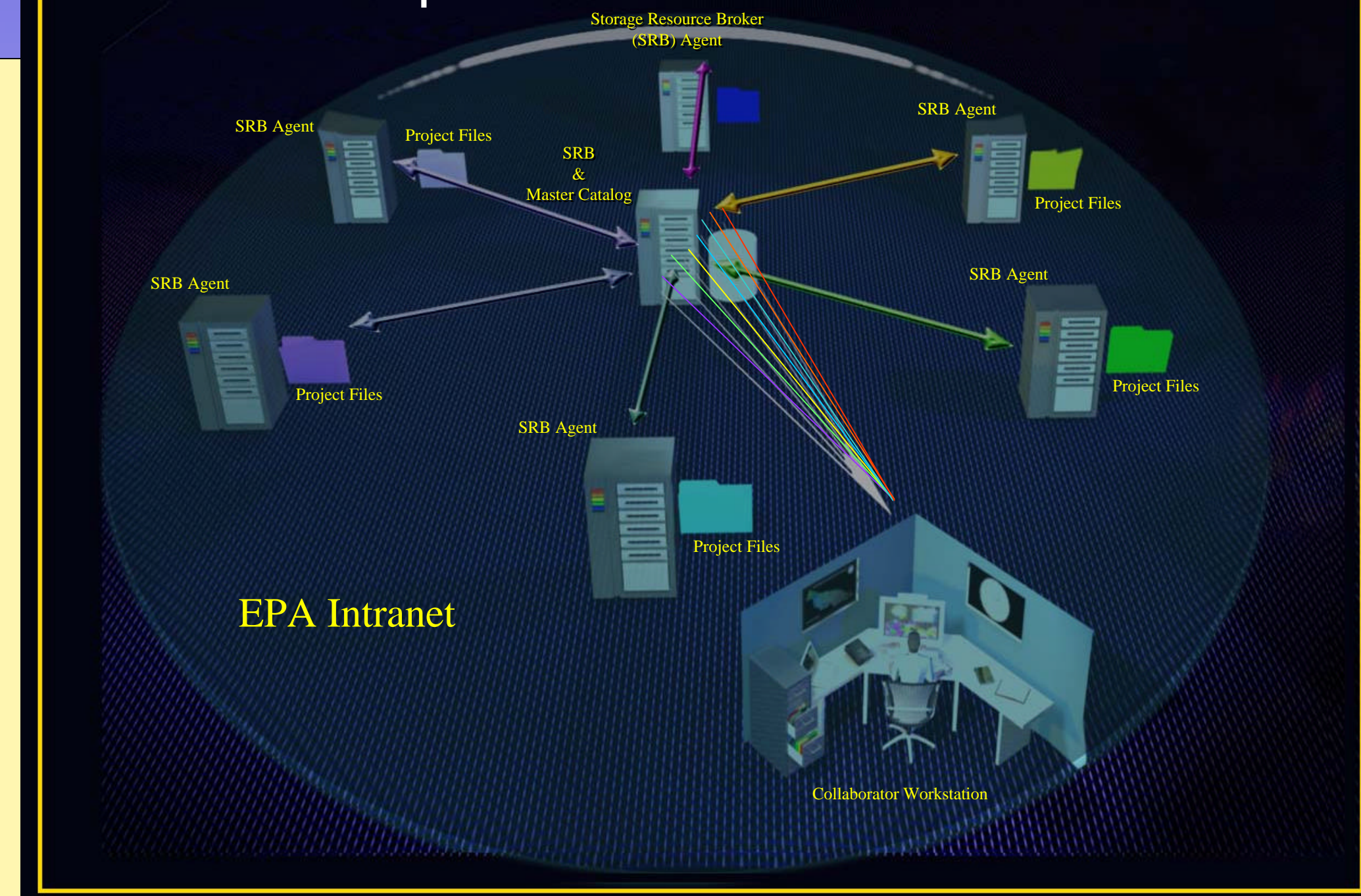
## GrADS Data Server (GDS)

- tool for analyzing and viewing geospatially-referenced data residing in geographically-distributed locations.
- Implemented in the EPA Intranet environment. Implementation on the Internet for publishing EPA data expected later this year.
- Uses GrADS - Grid Analysis and Display System - developed by the Center for Ocean-Land-Atmosphere Studies (COLA), Calverton, MD.
- Enhanced by EPA to support advanced search and visualization.
- Enables metadata search, subsetting of data, extraction, analysis, and visualization, including animation.
- Subsetting, extraction and transfer of data accomplished via the Open source Project for Network Data Access Protocol (OPeNDAP).
- Accessible from either a Windows or Linux machine. Only software required is a web browser.
- Currently, EPA's GDS provides access to the following dataset directories:
  - NOAA Operational Data Model Archive and Distribution System (NOMADS) Project;
  - NOAA Ocean Data Assimilation Experiments;
  - Precipitation/temperature data from NASA Goddard Space Flight Center;
  - Ocean Anomalies, GFS Forecasts, Global Landcover Classifications, and other associated data from COLA; and
  - EPA Air Quality data (CMAQ, two sets of data for 2001) and a small number of MODIS datasets retrieved from NASA.
- Plans are in place to provide access to additional datasets.

## GrADS Data Server (GDS)



## Proposed Data Grid Pilot



## Proposed Data Grid Pilot

- Goal: Provide an infrastructure for effective data sharing and collaboration among research team members at multiple dispersed locations.
- Status:
  - Preliminary plans in place.
  - Intend to use Storage Resource Broker (SRB) -- developed by University of California, San Diego -- for the pilot.
  - SRB provides single system image of all collaborator files on the Grid and supports data access.
  - Specific application for the pilot yet to be identified. Likely candidate: ORD/NERL's Remote Sensing Information Gateway.
  - Expect to complete prototype by end of FY06.

*Pioneering the use of Grid technology to leverage the aggregated intellectual and computational resources of EPA and partner organizations.*



**epascienceforum**  
Your Health • Your Environment • Your Future